Hyperarticulation as a signal of stance

Valerie Freeman [a]*

*Corresponding author: valerief@uw.edu, Tel: +1 208-596-0498, Fax: +1 206-685-7978
[a] University of Washington, Box 352425, Seattle, WA 98195-2425, USA

Abstract
This study analyzes an episode of a televised political talk show for evidence that speakers hyperarticulate concepts about which they express stances, a use of hyperarticulation that interacts with the discourse function of signaling new information. Using content analysis, utterances were coded on two dimensions: Evaluation (presence or absence of stance-expression) and Novelty (new or given information). To compare the resulting groups, four measures indicating hyperarticulation were used: speech rate of phrases, and the duration, pitch, and vowel space expansion (first and second formant values) of stressed vowels in the phrases. Group results showed significant effects for both Evaluation and Novelty, and an interaction between them. Stance-expressing items were hyperarticulated compared to a control group of neutral phrases, and within each group, new information was hyperarticulated compared to given information. Speech rate showed these effects most reliably, with vowel duration showing effects for Evaluation. Vowel space expansion showed the same patterns without statistical significance; pitch was not a reliable indicator. These findings provide acoustic correlates to stance-expression, which have not been extensively investigated previously and which can be applied in future work on the identification of specific types of stance.

Keywords
Hyperarticulation; stance; evaluation; discourse-given; discourse-new.

1.0 Introduction and Background

While a number of studies have considered the effect of emotions or moods on articulation (see Caelen-Haumont & Zei Pollermann 2008 for a summary), work attempting to identify acoustic correlates of stance-expression has found it difficult to employ systematic phonetic measurements, relying on impressionistic treatments of intonation or speakers' own characterizations (cf. discussions in Local & Walker 2008; Uldall 1960; Wichmann 2002a, 2002b). This study begins to fill this gap by examining an episode of the political television talk show *Tucker* for evidence that speakers use hyperarticulation to signal their stances about concepts they are discussing. Furthermore, this study proposes that by using hyperarticulation to convey their stances, speakers override the discourse convention of reducing the pronunciation of given information.

1.1 New vs. given

The *new* vs. *given* distinction as a discourse function relies on Grice's (1975) Cooperative Principle, the social agreement under which speakers try – and are expected – to give true, concise, and relevant information. As a practical component, Clark & Haviland (1977:4) added a "given-new contract" under which a speaker "agrees to convey information he thinks the listener already knows as given information and to convey information he thinks the listener doesn't yet know as new information."

Since "new" and "given" are relative terms, it is important to define them carefully. Prince (1981) offers a detailed taxonomy from which the broadest divisions are used for this study. Simply put, an entity is *new* when it is first introduced in a discourse segment, or "put on the counter" (p. 235), or when it is reintroduced after a change in the topic of conversation. *Given* information is everything already on the counter, or what the speaker assumes is in the listener's mental discourse model, whether overtly mentioned previously or inferable from logic or context.

There are many ways to signal the given-new distinction, including hyperarticulation and reduction. Many studies (e.g., Aylett & Turk 2004; Baker & Bradlow 2009; Bell et al. 2009; Chafe 1974; Fowler & Housum 1987; Jurafsky et al. 2001; Tomita 2008) have shown that items that are repeated, predictable, or familiar in discourse context have shorter durations, more contracted vowels, and weaker stress – in other words, given information is reduced (hypoarticulated) without sacrificing listener comprehension. Some (e.g., Aylett & Turk 2004; Baker & Bradlow 2009; Bolinger 1963; Jurafsky et al. 2001; Katz & Selkirk 2011; Tomita 2007) have also found that infrequent and unpredictable items have longer durations, greater pitch protrusion, and expanded vowel spaces – in short, new information is hyperarticulated.

1.2 Hyperarticulation

The goal of hyperarticulating is to "exaggerate speech sounds" (Whalen et al. 2004:155) or make phonetic and phonological components more distinct (de Jong, Beckman, & Edwards 1993). This is accomplished in several ways. Lengthening is a prevalent factor (Aylett & Turk 2004; Soltau & Waibel 2000; Wennerstrom 2001), and expanded vowel space also plays a key role (Tomita 2007). Hyperarticulated items receive heavier stress (Aylett 2005; de Jong et al. 1993), undergo changes in pitch contour and fundamental frequency (Soltau & Waibel 2000), experience less coarticulation, and have more precise places of articulation (de Jong et al. 1993).

While some studies have specified hyperarticulation as a means for signaling new information (e.g., Aylett 2005; Aylett & Turk 2004; Baker & Bradlow 2009; Katz & Selkirk 2011), several other uses are commonly discussed, including avoiding perceptual confusions (Whalen et al. 2004), correcting misunderstandings (Curl 2005; Soltau & Waibel 2000), emphasizing contrasts (Chafe 1974; de Jong et al. 1993; Katz & Selkirk 2011), and lending focus (Aylett 2005) or signaling something important (de Jong et al. 1993). Assuming speakers consider expressions of stance important, this study proposes that hyperarticulation can signal stance as well.

1.3 Evaluation and stance

Across disciplines, various terms are used to describe stances and their expression. In discourse analysis, *stance* may be used as an umbrella term for "personal feelings, attitudes, value judgments, or assessments" (Biber et al. 1999:966). This study is concerned with the more specific *attitudinal stance*, "speakers' subjective attitudes toward something" (Haddington 2004:101) and its expression, referred to as the social activity of *stancetaking*, or more broadly, *evaluation*, "perhaps the most salient and widely recognized form of stancetaking" (Du Bois 2007:142). Evaluation is "the broad cover term for the expression of the speaker or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about" (Hunston & Thompson 2000:5). This paper uses "Evaluation" to label the dimension of verbal stance-expression, identifying the presence or absence of stance without comparing specific subtypes.

In summary, cooperative speakers try to highlight new information by various means, including hyperarticulation, which may also be used to call attention to speakers' stances. Following Grice's Cooperative Principle, and with the given-new contract in mind, a cooperative speaker's hyperarticulation of given information as well as new may be interpreted not as a violation of the contract but as a signal of some other meaning the speaker intends the listener to grasp. If this is so, an interaction can be expected between the effects of hyperarticulation used to signal Novelty and those used to express stances.

2.0 Methods

Four acoustic measures that quantify hyperarticulation (speech rate, vowel duration, pitch, and vowel space expansion) were used to examine the interaction of two discourse-functional dimensions: Novelty (whether a phrase is new or given (repeated) information in a conversation) and Evaluation (whether the speaker expresses a stance about a concept). Utterances taken from conversational segments of a televised political talk show were divided on the Evaluation dimension into *stance* and *control* groups and on the Novelty dimension into *new* and *given*.

2.1 Program, segments, and speakers

To choose the talk show for this study, five episodes were randomly selected from a corpus of the audio tracks of televised political talk shows (Linguistic Data Consortium 2009). Since conversation was the context of interest, two episodes were eliminated due to their large proportion of non-conversational reporting. From the remaining three, one was randomly selected: an episode of MSNBC's *Tucker*, dated April 10, 2007 (Geist 2007). All segments (defined as broadcast between commercials) which consisted of conversations with at least one guest were analyzed – reports, announcements, introductions, and video clips were excluded. These conversational segments ranged in length from three to ten minutes, and none contained highly emotional speech, such as shouting, which would have been excluded, since emotional expression can affect articulation (for a summary of work on vocal cues to emotion, see Caelen-Haumont & Zei Pollermann 2008).

Six conversational segments with a total of five male speakers were analyzed. Two of these were interviews approximately four minutes long between the host, Tucker Carlson (age 37, from California) and a guest. One guest was the executive director of MoveOn.org, Eli Pariser (age 26, from Maine), who discussed his organization's hosting of Democratic primary debates, and the other guest was the chairman of the Center for the Rule of Law and Dean Emeritus of Boston University Law School, Ron Cass (approximate age: 50s, from Virginia), who discussed legal issues surrounding the paternity of the late Anna Nicole Smith's child. The other four segments included the host (again, Tucker Carlson) and two regular contributors to the show: conservative political commentator Pat Buchanan (age 68, from Washington, DC) and *Washington Post* columnist Eugene Robinson (age 52, from South Carolina). Because all participants called each other by first name during the program, they are referred to by first name here as well.

2.2 Selecting concepts and tokens

All content words (i.e., not function words) and phrases containing content words that were repeated at least three times by the same speaker in conversational portions of a segment were selected for analysis. Each repetition of lexically identical material (a word or phrase) is a *token*. A group of lexically identical tokens and all references to them (e.g., pronouns, synonyms, etc.) are taken together to form a *concept*. The following example, illustrated in Table 1, clarifies the terminology. In a four-minute interview, one guest referred to "the war in Iraq" more than three times, so the *concept* of "the war in Iraq" was selected for analysis. Table 1 displays some examples of the references he made to this concept, including truncations ("the war") and pronouns ("it, this"). The only portion common to at least three references was the word "war," repeated five times. Thus, each repetition of the word "war" was a token to be measured. However, not every repetition of the word "war" would necessarily pertain to the concept of "the war in Iraq." References to "the war in Afghanistan, the war on terror, culture wars," etc. would form their own, separate concepts, ensuring that speakers' stances about different concepts were not lumped together.

Table 1. *Relationship between concepts, references, and tokens*. References and tokens are components of a concept. All were considered together when coding Evaluation, and the resulting code (*stance* or *control*) applied to the whole concept, but each token was coded separately for Novelty (*new/given*).

Concept: "the war in Iraq"	
References	**Tokens**
"the war in Iraq"	war 1
"the war in Iraq"	war 2
"the war"	war 3
"a war"	war 4
"it"	
"this"	
"the war"	war 5
"this critical issue of Iraq"	

Inflectional variations were grouped together and considered a single concept, as long as the endings did not shift the word stress or cause a significant difference in meaning. For example, "defend, defends, defending" were all considered repetitions of the same concept, but only the portions common to all variants ("defend") were measured as tokens.

All references to a concept were considered together when coding for Evaluation, so the resulting code (*stance* or *control*) applied to the whole concept and every member token. The content analysis described in section 2.3 below was used to separate concepts into stance-expressing and control groups, making it possible to address the central proposal of this study, that speakers continue to hyperarticulate concepts about which they express stances, rather than reducing repeated material as expected for given information (as they should do for neutral/control concepts). To be able to compare *stance* and *control* concepts on their treatment of new and given information, each token was coded separately as *new* or *given*, as described in section 2.4.

2.3 Marking Evaluation (stance vs. control)

Content analysis was performed to identify speakers' expressions of stance regarding each concept. The coding scheme was formed by drawing on several analytical approaches to stance, as referenced in each category below. All references a speaker made to the concept, including truncations, abbreviations, ellipses, and pronouns, were considered evidence. One point was assigned to the concept for each stance-expressing action performed from the categories. Each category contains several actions, so a concept could receive several points per category, and a phrase used as evidence could perform more than one action, possibly in more than one category. The following describes each category together with examples from the talk show:

- **A. Introduction and repetition**: Speaker works to keep the topic in play by introducing or returning to it, overtly calling attention to it, repeating it when interrupted, or repeating similar references to it in close succession (cf. Labov 1972; Prince 1981). Examples of (re)introduction: "Look/Listen," "Let's talk about…," "Let me say this," and repetition: "A foolish comment, a nasty comment, an ugly comment."
- **B. Overt evaluation**: Speaker takes a stance by stating an opinion or making a prediction, including what "probably" happened or will happen (cf. Conrad & Biber 2000; Du Bois 2007; Hunston & Thompson 2000). Note that this is often called *subjectivity* (cf. Conrad & Biber 2000). Examples: "In my view," "The way I see it," "In all likelihood."
- **C. Evaluative description**: Speaker uses evaluative modifiers or commentary with a token or when referring to the concept (cf. Hunston & Thompson 2000; Labov 1972). Examples of modifiers: "ridiculous, important, impressive," and comments: "It turned my stomach."
- **D. Credibility**: Speaker offers support for a stance by expressing certainty, citing experts, personal credentials or experience, or by presenting the stance as fact (cf. Biber & Finegan 1989; Conrad & Biber 2000; Hunston & Thompson 2000). Note that this category covers *epistemic stance*, often treated separately from the *attitudinal stance* of the other categories here (cf. Conrad & Biber 2000). Speakers used these as evidence that their opinions were right. Examples: "That's a fact," "I was there," "Polls show…"
- **E. Persuasion and recommendation**: Speaker attempts persuasion, makes a suggestion or recommendation (cf. Conrad & Biber 2000). Examples: "Think of it this way," "They should be shut down."
- **F. Agreement**: Speaker agrees or disagrees with another speaker (cf. Conrad & Biber 2000; Du Bois 2007). Examples: "I (dis)agree," "Absolutely," "Not at all."

To calculate a score for the concept, the number of items of supporting evidence was divided by the number of tokens (repetitions of the concept word/phrase). If the ratio was 2.00 or higher, the concept was marked *stance*; otherwise, it was marked *control*. This ratio was determined by creating a frequency distribution of all concept scores for all speakers. The distribution was nearly normal, with a mean of 1.92, divided approximately in half by the 2.00 cutoff point (see Table 2). (This binary treatment was verified with post-hoc analysis that showed no correlation between raw concept score and any of the dependent variables.)

The following orthographically transcribed excerpt illustrates how one concept was coded. In it, Eli talks with Tucker, the host, about his organization's call to withdraw U.S. troops from Iraq.

> Eli: Um and and actually, while we're on that topic, let's talk about terminology for a second because, you know, I call[ed] it a war, but really it's an occupation. And and, this is one of the key points that I think uh people obscure. You know, this is an occupation, and the question, you don't win or lose an occupation. Uh you just, it's a question of how quickly uh you you remove your troops.

Tucker: Right, I I guess uh, I mean occupation is of course a loaded term itself because it implies the presence is illegitimate or or part of a a colonial rule.

Eli: Well no, it's just, it's saying our our troops are in someone else's country and they don't want us there, and that's that's a fact. I mean if you ask most Iraqis, they at this point uh, you know, they don't think that we're helping. They think that we're hurting. And if you ask most Americans, they think the same thing.

Eli introduces the concept of "calling the war in Iraq an occupation" and overtly calls attention to his intention to do so by saying, "let's talk about terminology." So, this concept (represented by his tokens of "occupation") received two points under Category A: one for the introduction and one for overtly calling attention to his desire to discuss it. It also received a point in Category B for the overt opinion ("I think…") and one in Category C for the descriptor "key points" because "one of the key points" refers to calling the war an occupation. It received another point in Category C for the descriptive comment "it's saying our our troops are in someone else's country and they don't want us there." Three points were awarded under Category D for "that's a fact," "if you ask most Iraqis," and "if you ask most Americans, they think the same thing." Finally, one point was given under Category F for his disagreement with Tucker's characterization of his use of the term occupation ("well no… it's saying…"). In total, the concept received 9 points, which were divided by Eli's 3 repetitions of "occupation" for a score of 3.00; with a score over 2.00, the concept and all tokens of "occupation" were coded as *stance* rather than *control*.

2.4 Marking Novelty (new vs. given)

Concepts were selected for analysis when repeated three or more times by the same speaker within a conversational portion of a segment, and each repetition (token) was coded as either *new* or *given.* Following Prince's (1981) taxonomy, the first utterance of a concept in a segment was coded as *new*, even if said in a non-conversational portion or by another speaker. This indicated the instance that introduced the concept into the discourse, making it fresh in all participants' minds. If the first utterance of a concept for one speaker was said in a non-conversational portion and/or by another speaker, it was not measurable as a token. For example, in one segment, the host, Tucker Carlson, introduced the topic to be discussed by reporting on a Democratic primary debate. He said the word "debate" twice in his introduction and again in his first question to the other commentators, but never again in that segment. Although he said "debate" three times, only once was during a conversational portion, so *debate* was not used as a concept for Tucker. However, in a response to Tucker's question, the commentator Eugene Robinson repeated "debate" three times, making *debate* an analyzable concept for him. Because it was Tucker who introduced *debate* into the discourse, none of Eugene's repetitions of the concept could be coded as *new*.

If a concept was dropped in the discussion but later picked up again, the reintroduction was coded as *new*. Conversational segments lasted 3-10 minutes and regularly shifted focus in under a minute, so a concept was considered reintroduced when it followed at least five speaker-turns which spanned 60 seconds or more and did not include any reference to the concept in question. This ensured that the concept was no longer "on the counter" in the discourse (Prince 1981), and a reintroduction was necessary to bring it back to the forefront in speakers' minds. Finally, all tokens not coded as *new* were marked *given*. With concepts (and their member tokens) coded as *stance* or *control* and each token as *new* or *given,* the two categories of labels combined to yield four types of tokens: *new-stance, new-control, given-stance,* and *given-control*.

2.5 Concepts, tokens, and vowels

The coding procedures identified a total of 65 concepts containing 218 tokens (repetitions of the concept word/phrase), evenly divided between *stance* and *control*. Because some token phrases contained more than one stressed vowel, the total number of vowels measured was higher, 237. Table 2 shows the number of concepts, tokens, and vowels with each code combination, including breakdowns for *new* and *given*. Overall, the sample is balanced: there are similar numbers of *stance* and *control* items in each category, and there are two to three times as many *given* as *new* tokens, which is expected since a token is likely to be repeated more frequently as given information than to be introduced as new.

Table 2. ***Number of concepts, tokens and vowels by code.***

Type	Concepts	Tokens			Vowels [a]		
		Given	*New*	Total	*Given*	*New*	Total
Control	33	82	27	**109**	94	31	**125**
Stance	32	73	36	**109**	75	37	**112**
Total	**65**	**155**	**63**	**218**	**169**	**68**	**237**

[a] Some token phrases had multiple measurable (stressed) vowels.

The sample was also balanced for other factors known to affect hyperarticulation, including intrinsic vowel properties, token length, and lexical frequency. Cross-linguistically, low vowels generally have longer durations and lower pitches than high vowels (cf. e.g., Keating 1985; Whalen & Levitt 1995). Vowels in the sample were distributed fairly evenly by height, with the ratios of codes within each height reflecting those in Table 2. In addition, English lax vowels are generally shorter than their tense counterparts (cf. Keating 1985); the sample was also balanced for tenseness, with approximately equal numbers of tense and lax vowels, again distributed reflecting the ratios in Table 2. Longer words generally have shorter syllables, so it was important to balance the sample for token length to avoid confounding the durational measures. There were roughly equal numbers of monosyllabic, disyllabic, and polysyllabic (3-6 syllables) tokens, and the codes for tokens of each length reflected the ratios in Table 2. Many studies (e.g., Baker & Bradlow 2009; Bell et al. 2009; Jurafsky et al. 2001) have found that more frequent words are more reduced as reflected in measures such as word and segment durations, vowel reduction, and segment deletion. Lexical frequencies for this study were determined using the online interface of the Corpus of Contemporary American English (COCA, Davies 2012). COCA is an actively growing corpus comprising about 20 million words per year since 1990, drawn equally from five genres: transcripts of conversational TV programs such as the talk show used in this study, newspapers, magazines, fiction, and academic journals. Because the talk show examined here aired in April 2007, only material up to 2006 was searched, a subset of about 352 million words. Frequencies were based on lemma and part of speech in order to include inflectional endings but not compounds. As with the other factors, the sample was balanced for lexical frequency. About 80% of the tokens were distributed fairly evenly across lexical frequencies up to 500 per million (pm), with the rest spreading up to 4100 pm. Within each 100-pm frequency range (with all above 500 pm grouped together), the distribution of codes reflected the ratios shown in Table 2, with the exception of very low frequency words (less than 50 pm), in which more *given* tokens were marked as *stance* (16 vowels) than *control* (4 vowels). (A post-hoc analysis also showed no correlation between word frequency and any of the dependent variables.)

Finally, each speaker is a potential source of variation on any acoustic dimension. Although the totals for all speakers combined are balanced, not all speakers contributed to each cell equally. The host, Tucker, contributed the largest proportion of *control* tokens (44% of *new-control*, 38% of *given-control*) but the smallest proportion of *stance* tokens (6% each). Pat and Ron varied, each contributing between 15% and 50% of each cell, while Eli was fairly constant at 11%-16%. Eugene contributed 8%-9% of each cell except *new-control*, where he had no tokens, resulting in the removal of his data from measures comparing all four cells.

2.6 Measurements

One goal of this study was to establish that speakers' stances are signaled by measurable acoustic correlates. With the hypothesis that stance-expressing utterances are hyperarticulated, four measures indicating hyperarticulation were chosen: speech rate over a word/phrase, and the duration, pitch (f_0), and first and second formant values (F1, F2) of stressed vowels within the word or phrase. Speech rate and duration are measures of lengthening, a prominent feature of hyperarticulated speech. Normalized pitch difference, the amount a pitch deviates from a speaker's mean pitch, can be used to detect whether pitch is responsive to stance-expression. Vowel space is often used to define the continuum between reduced (contracted) and hyperarticulated (expanded). Stressed vowels were chosen as the focus for measurement because they are targets for lengthening in English, they often carry the most prominent pitch in a word, and changes in their formant values more readily reflect degrees of reduction as compared to unstressed vowels, which tend to be mid-centralized in F1xF2 vowel space.

Before vowels were measured, their onsets and offsets were manually demarked on text tiers in *Praat* (Boersma & Weenink 2008). The onsets of vowels after stops were marked just after the consonant release burst, as indicated by increased energy throughout the range of the spectrogram and/or a short period of increased amplitude in the waveform. Vowel onsets following voiceless stops were marked similarly so that the vowel duration included aspiration. The offsets of vowels before stops were marked at the consonant closure, as indicated by the loss of energy of F2 and/or a drop in amplitude in the waveform. Vowels adjacent to nasals were marked just inside the depression in F2 that indicates the nasal consonant, coinciding with a less complex waveform. Vowels neighboring fricatives were marked up to but not including frication noise, as indicated by energy in high frequencies and a dense, medium-amplitude waveform. Onsets and offsets of vowels adjacent to glides and liquids were determined by changes in waveform in combination with changes in F2 and/or F3. For the rhotic /ɹ/, the vowel boundary was characterized by a lower-amplitude, less complex waveform and a dip in F3. To minimize effects of ɹ-coarticulation, vowels adjacent to rhotics were marked at a point where the slope of F3 flattened, coinciding with lower amplitude in the waveform. If necessary, the waveform was magnified and visually inspected to identify the change in its complexity. For the lateral /l/, the vowel boundary was characterized by a rise in F3. As with the rhotic, it was often necessary to rely on changes in the complexity and amplitude of the waveform to identify vowel onsets and offsets. Glides were characterized by changes in F2: a rise for /j/ and a fall for /w/. The offsets of diphthongs were marked at the extreme point of the glide, which often coincided with a change in waveform amplitude, as were vowels following or preceding consonantal glides.

The duration, midpoint formants (F1, F2), and pitch (f_0) of stressed vowels were measured using a script in *Praat* that referred to manually demarked vowels. The script paused before and after measuring each vowel to allow manual verification that pitch and formant tracking were accurate, as well as the resulting measurements. The formant range was set to 0-5500 Hz with a window length of 25 ms, dynamic range of 50 dB, and 14 formant coefficients (6 formants), with the pitch range set to 60-500 Hz.

2.6.1 Lengthening measures

Lengthening was examined by measuring the speech rate (syllables/second) of all 218 tokens and the duration (ms) of all 237 stressed vowels. Speech rate was calculated by dividing the length in syllables of each token's canonical form by its measured duration in seconds. Both measures were expected to show lengthening, with a slower rate and longer stressed vowel durations for new information and *stance* tokens as compared to given information and *control* tokens. Since an interaction was expected between Novelty and Evaluation, the two dimensions were examined separately and in combination. Thus, the means of *new* and *given* tokens were compared, followed by *stance* and *control*, and finally combinations of the two (*given-control, new-control, given-stance, new-stance*).

2.6.2 Pitch measures

This study used normalized pitch difference – the amount a pitch deviates from a speaker's mean pitch – because pitch deviations *in either direction* can draw listener attention and signal word prominence. Differences were expected to be larger for highlighted words/phrases (new information and/or *stance* tokens as compared to given information and *control* tokens). Nine vowels were excluded from this measure because background noise or another speaker's voice made pitch unmeasurable, leaving 228 stressed vowels measured from 210 tokens. Pitch (f_0) was measured in Hz at the midpoint of stressed vowels using *Praat*'s autocorrelation algorithm with a pitch range of 60-500 Hz. Each speaker's mean pitch was calculated as the mean of all pitch values at midpoint for all measured stressed vowels. Because each speaker had a different mean pitch and range, values were normalized using within-speaker z-scores, each of which represented a midpoint pitch's deviation from the speaker's mean. Z-scores were then averaged for each combination of token type (*new/given, stance/control*) and compared for all speakers combined and for each speaker separately.

2.6.3 Formant measures

To investigate the effects of Novelty, Evaluation, and their interaction, only vowel phonemes which had tokens of all four combinations (*new-stance, given-stance, new-control, given-control*) said by the same speaker were used. This reduced the number of vowels examined to 62, with a small number of *stance* tokens and only one or two phonemes per speaker. First and second formants (F1, F2) were measured in Hertz at the midpoint of stressed vowels. Within each speaker, the F1xF2 values for each type combination were averaged for each vowel. To quantify the effects of Novelty and Evaluation, the Euclidean distances between the means of each type combination were calculated and compared. The relationships are illustrated in Figure 1, a conceptual diagram of an F1xF2 vowel space. Nodes represent mean values of tokens with each Novelty-Evaluation coding combination; lines indicate the Euclidean distances that represent the effect of one dimension (Novelty or Evaluation) on tokens of each category of the other dimension. For example, the black dashed line Nov(stan) shows the distance between *given-stance* (GS) and *new-stance* (NS) tokens and represents the effect of Novelty on *stance* tokens; the solid black line Eval(new) shows the distance between *new-stance* (NS) and *new-control* (NC) tokens, representing the effect of Evaluation on *new* tokens. This diagram depicts the hypothesized interactions between Novelty and Evaluation: Evaluation will have a greater effect, as seen by the longer lengths of the solid Eval lines compared to the dashed Nov lines, and there will be an interaction such that Novelty affects *stance* tokens more than *control* and Evaluation affects *new* more than *given*.

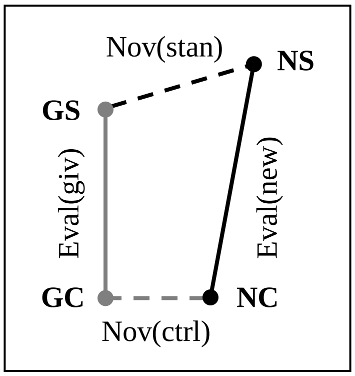

Figure 1. *Conceptual diagram of effects of Novelty and Evaluation*. Nodes represent mean F1xF2 values for vowels of each coding combination; N = *new*, G = *given*, S = *stance*, C = *control*. Lines represent effects of the dimensions on types of tokens, e.g., Eval(new), the distance between *new-stance* and *new-control* tokens, represents the effect of Evaluation on *new* tokens.

3.0 Analysis

3.1 Hypotheses

This study aims to determine how Novelty and Evaluation interact as effects on hyperarticulation, as measured by speech rate, vowel duration, normalized pitch difference, and formant values of stressed vowels. Three hypotheses were tested for each measure:

H1: There is a reliable effect for Novelty. As in previous studies (e.g., Aylett 2005; Aylett & Turk 2004; Baker & Bradlow 2009; Katz & Selkirk 2011), speakers are expected to hyperarticulate new information. Speech rate should be slower, stressed vowel durations longer, and pitch differences and vowel space expansion greater for *new* than *given* information.

H2: There is a reliable effect for Evaluation. Speakers are expected to hyperarticulate words and phrases about which they express stances (*stance* tokens) relative to their articulation of phrases about which they do not express stances (*control* tokens). Speech rate should be slower, stressed vowel durations longer, and pitch differences and vowel space expansion greater for *stance* than *control* tokens.

H3: There is an interaction between Novelty and Evaluation. The interaction may be additive, or some types of tokens may be affected more than others by one or both dimensions. It is predicted that Evaluation will have a greater effect than Novelty overall, such that all *stance* tokens will be hyperarticulated compared to all of their *control* counterparts. Individual variation is also expected in that speakers may employ each measure to different degrees or in different directions.

If all three hypotheses are supported, many different interactions are possible. Each combination of *new/given* and *stance/control* tokens might be articulated to a different degree. For example, Novelty may have a greater effect for *stance* than *control* tokens, or Evaluation may affect *new* more than *given* tokens, or both, resulting in a combination where *new-stance* tokens are hyperarticulated more than any other type of token, as depicted in Figure 1.

3.2 Statistics

Factorial ANOVAs, linear mixed effects analysis, and t-tests were performed using *R* (R Development Core Team 2008). A separate ANOVA was performed for each of three dependent variables: Speech Rate (syllables/sec), Vowel Duration (ms), and z-score normalized Pitch Difference. In all three, the independent variables were Evaluation (stance, control), Novelty (new, given), and Speaker. For Speech Rate and Vowel Duration, linear mixed effects models were employed with Novelty and Evaluation as fixed effects and Speaker as a random effect. In an ANOVA for formant measures, the dependent variable was the Euclidean Distance (Hz) between token types. With this measure, Novelty and Evaluation affected non-overlapping subsets of token types, so they were entered into the ANOVA as one independent variable with two levels: Dimension (Novelty, Evaluation), and Speaker was a second factor. Additionally, a post-hoc t-test was performed for each pair of Euclidean distances using an alpha of 0.05 and a Bonferroni Dunn corrected alpha of 0.0083.

4.0 Results

4.1 Speech rate and stressed vowel duration

Evidence from mean speech rates and stressed vowel durations showed effects for both Novelty and Evaluation. A three-way factorial ANOVA for Speech Rate showed significant effects for all three independent variables, Evaluation ($F(1, 218) = 20.66$, $p < 0.001$), Novelty ($F(1, 218) = 8.72$, $p = 0.004$), and Speaker ($F(4, 218) = 11.82$, $p < 0.001$), and for the interaction between Evaluation and Speaker ($F(4, 218) = 8.34$, $p < 0.001$). For Vowel Duration, significant effects were Evaluation ($F(1, 218) = 7.88$, $p = 0.005$) and Speaker ($F(4, 218) = 6.87$, $p < 0.001$), with an interaction between them ($F(4, 218) = 6.59$, $p < 0.001$). For both measures of lengthening, *stance* tokens are said more slowly than *control* tokens, and *new* more slowly than *given*, although the difference between *new* and *given* is not significant for Vowel Duration.

To better account for the effect of Speaker, this effect was entered as a random factor into linear mixed effects models with Novelty and Evaluation remaining as independent variables. For both dependent variables, Speech Rate and Vowel Duration, Evaluation was not found to have a significant effect, leaving only Novelty as a significant effect on Rate ($p < 0.01$). This suggests that for this sample, variation between speakers may confound the results for Evaluation.

With Novelty and Evaluation labels combined, a trend emerges suggesting an interaction between the two factors, but the interaction is not statistically significant for either measure. Figures 2-3 show each token type horizontally with mean speech rate (Figure 2) and stressed vowel duration (Figure 3) vertically. In both, Evaluation has a greater effect than Novelty: all *stance* tokens are said more slowly than all of their *control* counterparts. Novelty behaves in the expected direction (slower for *new* than *given* items), except for the vowel duration of *stance* tokens, in which *given* and *new* are not significantly different.

Individual behavior was examined but did not show uniform patterns; in two-way ANOVAs for each speaker, Evaluation and Novelty were significant factors only sporadically. For Rate, Evaluation was significant for Ron ($F(1, 52) = 53.92$, $p < 0.001$) and Pat ($F(1, 64) = 11.89$, $p = 0.001$), and Novelty for Tucker ($F(1, 50) = 4.52$, $p = 0.039$), with no interactions. For Duration, Evaluation was significant for Ron ($F(1, 52) = 7.67$, $p = 0.008$) and Eugene ($F(1, 18) = 10.46$, $p = 0.005$), with no interactions. Ron's patterns are in the predicted directions, reflecting the patterns in Figures 2-3, but contradictory patterns are also present: Pat's *stance* tokens were faster than his *control* tokens, Tucker's *new* faster than his *given*, and Eugene's *control* vowel durations were longer than his *stance* vowels. Finally, Tucker generally spoke more quickly and had shorter stressed vowels than the other speakers.

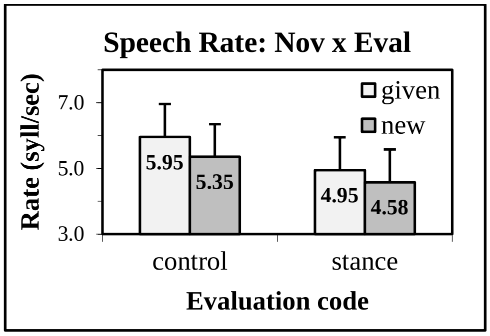

Figure 2. ***Mean speech rate, Novelty-Evaluation interaction*****.** Rate (syllables/second) shown on vertical axis, Evaluation code on horizontal, Novelty code by shading (*given* in light, *new* in dark); error bars indicate a confidence interval of 0.95.

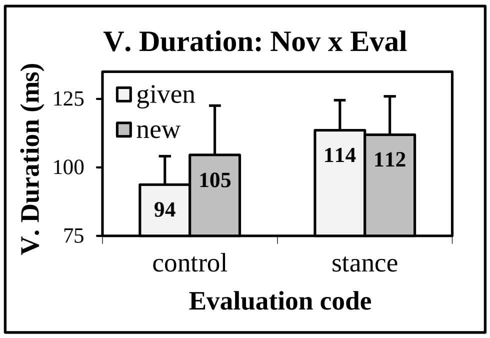

Figure 3. ***Mean stressed vowel duration, Novelty-Evaluation interaction*****.** Vowel Duration (ms) shown on vertical axis, Evaluation code on horizontal, Novelty code by shading (*given* in light, *new* in dark); error bars indicate a confidence interval of 0.95.

4.2 Normalized pitch difference

With all speakers combined, a three-way ANOVA showed no significant effect on the dependent variable normalized Pitch Difference for any independent variable – Novelty (new, given), Evaluation (stance, control), Speaker – or any interaction between them. As seen in Figure 4, there is no significant difference between *new* and *given* or between *stance* and *control*; the group pitch results offer no support for any of the three hypotheses.

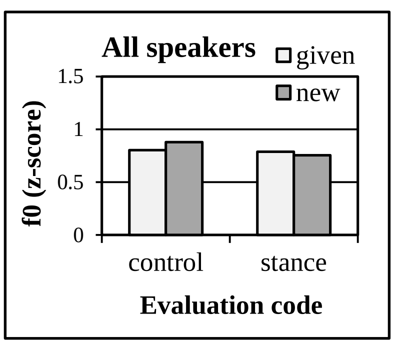

Figure 4. ***Normalized pitch difference, Novelty-Evaluation interaction, all speakers*****.** Mean z-scores shown on vertical axis, Evaluation code on horizontal, Novelty code by shading (*given* in light, *new* in dark).

Because there was a large amount of variation in how each speaker used pitch, within-speaker effects were examined. While ANOVAs for each speaker found no reliable effects, it appears that the relatively even bars shown in Figure 4 may result from differing individual behavior (Figure 5). Although these patterns can be described impressionistically, no reliable conclusions can be drawn about the relaionshp between pitch and Novelty or Evaluation without a more comprehensive intonational analysis than was possible for this paper.

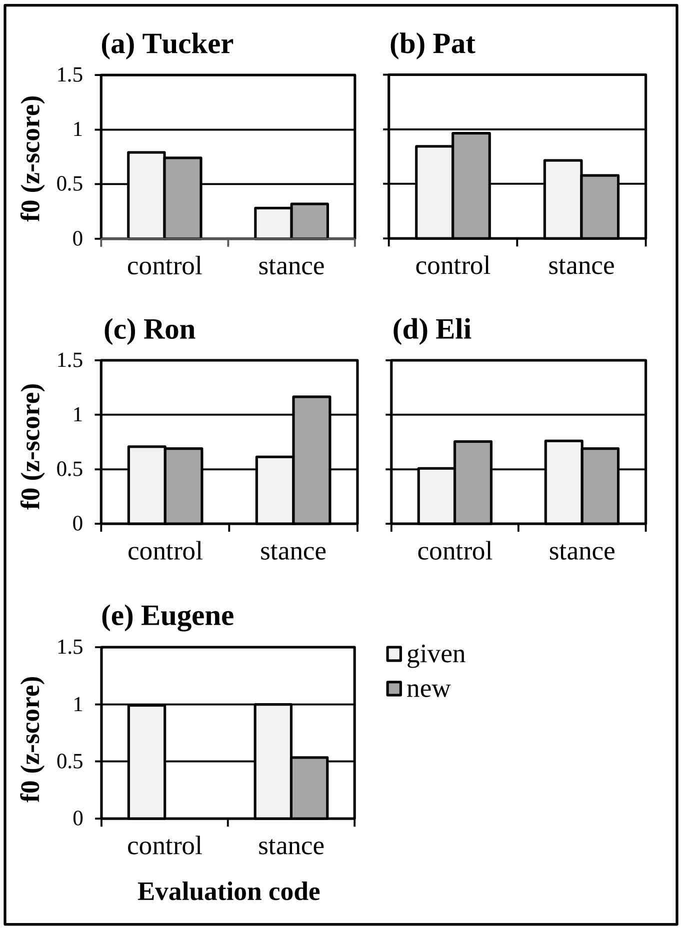

Figure 5. ***Normalized pitch difference, Novelty-Evaluation interaction, individual speakers*.** Mean z-scores for pitch shown on vertical axes, Evaluation code on horizontal, Novelty code by shading (*given* in light, *new* in dark).

4.3 Formant values

Vowel space measures suggest limited support for the hypotheses. For vowels with all four combinations of token types, a two-way ANOVA was performed to determine whether there was a difference between the effects of Novelty and Evaluation. The two effects were entered as one independent variable with two levels, Dimension (Novelty, Evaluation), and Speaker was entered as the second factor. Dimension was the only significant factor ($F(1, 20) = 10.35$, $p = 0.004$), with no interaction. To determine whether the effects of each dimension on each token type were significantly different, a post-hoc t-test was performed for each pair of Euclidean distances depicted in Figure 1 using an alpha-level of 0.05 and a Bonferroni Dunn corrected alpha-level of 0.0083. Although the distances are arranged in the expected pattern (Figure 6), with Evaluation having a greater effect than Novelty overall, Evaluation affecting *new* more than *given* tokens, and Novelty affecting *stance* more than *control* tokens, the only significant difference was between the largest and smallest distances, Evaluation(new) and Novelty(control) ($t = -3.407$, $df = 11.976$, $p = 0.005$).

As with the other measures, individual variation played a role, but with few data points per speaker, individual effects are not clear. There was no consistent pattern in the arrangement of token types (nodes in Figure 1), but for five of the seven vowel phonemes examined, the distances representing the effects of Evaluation were longer than those representing the effects of Novelty.

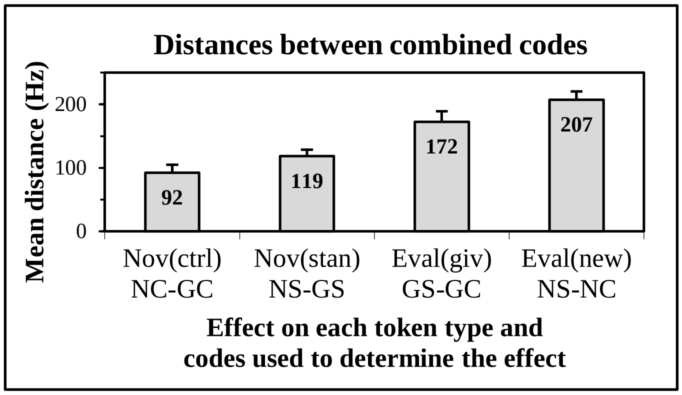

Figure 6. *Effects of Novelty and Evaluation on each type of token, all speakers*. Effects on token types shown on the horizontal axis, e.g., Nov(ctrl) = the effect of Novelty on *control* tokens, followed by the combinations of types used to determine the size of the effects, e.g., NC-GC = the distance between *new-control* and *given-control* tokens, measured in Hz on the vertical axis. Error bars indicate a confidence interval of 0.95. Evaluation and Novelty have significantly different effects ($p = 0.004$), but for individual effects, the only significant difference is between Nov(ctrl) and Eval(new) ($p = 0.005$).

5.0 Discussion

5.1 Conclusions

Some support was found for all three hypotheses:

H1: There is a reliable effect for Novelty. Speech rate demonstrated that new information was hyperarticulated compared to given information.

H2: There is an effect for Evaluation. Speech rate and stressed vowel duration showed that speakers hyperarticulated words and phrases about which they express stances (*stance* tokens) when compared to neutral phrases (*control* tokens). However, when Speaker is treated as a random factor in a linear mixed effects model, Evaluation does not have a significant effect.

H3: There is an interaction between Novelty and Evaluation. Speech rate showed an additive effect with Evaluation having a greater impact than Novelty, such that all *stance* tokens were hyperarticulated compared to all of their *control* counterparts, and within each condition, *new* tokens were hyperarticulated compared to *given*. Vowel formants also showed this pattern, but with limited statistical significance. Individual variation played a role in every measure but was particularly strong with normalized pitch difference, the only measure to show no reliable group result.

While the expected patterns were found, statistical support for the effect of Evaluation and its interaction with Novelty is not definitive. With a small sample size and high inter-speaker variation, the generalizability of results to a larger corpus is not yet clear. However, some of the phonetic measures used in this study may still prove useful as indicators of evaluative expression. Speech rate was the most reliable indicator of group behavior, showing the patterns predicted for all three hypotheses: speakers hyperarticulate new information, and in addition, they hyperarticulate both new and repeated material about which they express stances, even more than new information about which they express no stance. Vowel space expansion showed the same patterns, but less reliably and less clearly with the small sample size than might be found with more data. Stressed vowel duration also appears promising as an indicator of stance, but normalized pitch difference showed no reliable result, possibly due to greatly differing individual patterns.

5.2 Individual variation

In aggregating the patterns of individual variation found with each measure, it appears that speakers may employ different combinations of measures in their treatment of Novelty and Evaluation. For example, Ron behaves as predicted using nearly every measure, hyperarticulating *new* over *given* and *stance* over *control*, with Evaluation having a greater effect than Novelty overall. Pat is also fairly consistent in his use of Evaluation, but not in the expected direction: he tends to hyperarticulate *control* rather than *stance* tokens, or perhaps a more accurate conclusion is that he reduces rather than hyperarticulates *stance* tokens. This does not completely contradict the hypotheses made in this study; rather, Pat may habitually vary his articulations widely so that flattening various features is contrastive. These patterns are suggestive of differing strategies that could be modeled more reliably in a larger study. In this study, the interaction often found between speaker and Evaluation could be a reflection of the speakers' unequal contributions of token types; a larger data set could balance such contributions.

5.3 Future work
As some of the first work to identify acoustic correlates of stance-expression, the scope of this study was intentionally constrained. Many of its limitations would be improved with a larger, more controlled data set. An expansion of this study is underway, beginning by building a corpus of spontaneous speech recorded while pairs or small groups complete collaborative tasks designed to elicit a high density of stancetaking at varying levels of engagement. This will allow for systematic control of a variety of factors which will be tested for their effects on stance-expression, including speaker demographics, group composition, subject matter, and level of engagement with the topic or interlocutor. In the talk show used here, speakers were all male and spanned a range of ages and dialect regions. With a large, balanced sample, it may be possible to discover whether the individual variation found in this study is actually reflective of socially-differentiated patterns.

In addition, a larger project will enable more rigorous phonetic analysis. This study examined mainly stressed vowels, but consonant closures are also manipulated along various acoustic dimensions, including components of hyperarticulation such as duration, degree of stricture, and place of articulation. Useful information may also be expected throughout the vowel, rather than just at midpoint. This study focused on target words in repeated material, but stances may be expressed on other parts of utterances as well. Longer stretches of speech should be examined, either to provide a baseline for comparison with target words, or to identify areas of change that may indicate evaluative expression. Durational measures such as those used in this study can be extracted and compared in a straightforward manner, but other measures may need more careful treatment. In particular, pitch patterns have been found to correlate with so many different linguistic and social factors that a more comprehensive intonational analysis may be necessary to tease apart effects of evaluative expression. Potential confounds to the current pitch results may include such factors as tokens' utterance positions and the information- or discourse structure of those utterances. It is possible that an effect for evaluation could be found after accounting for patterns of prominence predicted by models of intonational semantics and pragmatics (e.g., Hirschberg 1993; Hirschberg & Pierrehumbert 1986; Pierrehumbert & Hirschberg 1990). In addition, this study compared only pitches at the midpoint of stressed vowels, but a more sophisticated treatment of intonation would include measurements over more points in target syllables and over entire intonational phrases. A larger data set would also provide the power to find reliable patterns in other multidimensional measures, such as vowel space expansion.

Finally, this study only had the power to consider evaluation at a broad level, but a larger project may also be able to investigate acoustic correlates of more specific types of evaluation, stance, or attitudes, including their polarity or strength. For example, the evaluation coding scheme in this study included epistemic stance (credibility, certainty) when used as support for attitudinal stance moves, but the expression of different types of stance may involve different acoustic correlates. Such an undertaking necessarily involves the continued integration of methods from multiple fields. Wichmann (2002b) recognized the difficulty of combining local phonetic measurements with more holistic approaches such as discourse- and conversation analysis, which rely on context, discourse structure, and subsequent speaker turns to interpret utterances. Although it may be challenging, researchers looking at similar questions from different perspectives can benefit from combining methods. This study is an example, employing methods of content analysis to identify areas of interest for subsequent phonetic measurement. Local & Walker (2008) and Curl (2005) provide other excellent examples of the integration of conversation analysis and phonetics. It is hoped that once phonetic correlates to discourse-level phenomena are identified, they can be found more easily in corpora, and both approaches can again be applied to further refine their categorizations.

Acknowledgements

Special thanks to Dr. Richard Wright, Dr. Betsy Evans, Dr. Daniel McCloy, Dr. Alicia Wassink, members of the UW Phonetics Lab, three anonymous reviewers and the editor, Dr. Kenneth de Jong, for their helpful comments on earlier versions of this work. Thanks to Dr. Emily Bender and Dr. Mari Ostendorf for their assistance with the corpus data.

This report is based on work conducted for a master's thesis filed at the University of Washington, August 2010 (Freeman 2010). Portions of this research were presented at the 160th Meeting of the Acoustical Society of America (ASA) in Cancun, Mexico, November 15-19, 2010, and at the 26th Northwest Linguistics Conference (NWLC) in Burnaby, BC, Canada, May 8-9, 2010.

Transcript and annotation data provided by the Linguistic Cues of Roles in Conversational Exchanges (LiCORICE) Project, funded by NSF grant IIS-0811210, and by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), Contract No. W911NF-09-C-0131. All statements of fact, opinion or conclusions contained herein are those of the author and should not be construed as representing the official views or policies of IARPA, the ODNI or the U.S. Government.

Reference List

Aylett, M. (2005). Synthesising hyperarticulation in unit selection TTS. *Proceedings of the European conference on speech communication and technology (Interspeech 2005)* (pp. 2521-2524).

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31-56.

Baker, R. E. & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style and prosody. *Language and Speech*, *52*(4), 391-413.

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language 60*(1), 92-111.

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text* - *Interdisciplinary Journal for the Study of Discourse, 9*(1), 93-124.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Longman.

Boersma, P., & Weenink, D. (2008). *Praat, a system for doing phonetics by computer* (version 5.0.34) [computer software]. Retrieved from http://www.praat.org/.

Bolinger, D. L. (1963). Length, vowel, juncture. *Linguistics*, *1*(1), 5-29.

Caelen-Haumont, G., & Zei Pollermann, B. (2008). Voice and affect in speech communication. In K. Izdebski (Ed.), *Emotions in the human voice, volume 1* (pp. 215-232). San Diego: Plural Publishing.

Chafe, W. L. (1974). Language and consciousness. *Language, 50*(1), 111-133.

Clark, H. H., & Haviland, S. (1977). Comprehension and the given-new contract. In R. Freedie (Ed.), *Discourse production and comprehension* (pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conrad, S., & Biber, D. (2000). Adverbial marking of stance in speech and writing. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 56-73). New York: Oxford University Press.

Curl, T. S. (2005). Practices in other-initiated repair resolution: The phonetic differentiation of 'repetitions.' *Discourse Processes, 39*(1), 1-43.

Davies, M. (2012). The corpus of contemporary American English: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

de Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech, 36*, 197-212.

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139-184). Amsterdam: Benjamins.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489-504.

Freeman, V. (2010). Using acoustic measures of hyperarticulation to quantify novelty and evaluation in a corpus of political talk shows. Master's thesis. Seattle: University of Washington Press.

Geist, W. (executive producer) (2007, April 10). Tucker [Television broadcast]. New York: MSNBC.

Grice, P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics, 3: Speech acts*. New York: Academic Press. Reprinted in Geirsson, H. & Losonsky, M. (Eds.) (1996). *Readings in language and mind* (pp. 121-133). Cambridge, MA: Blackwell Publishers.

Haddington, P. (2004). Stance taking in news interviews. *SKY Journal of Linguistics*, *17*, 101-142.

Hirschberg, J. (1993). Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence, 63*(1-2), 305-340.

Hirschberg, J., & Pierrehumbert, J. (1986). Intonational structuring of discourse. Proceedings of the 24th Meeting of the Association for Computational Linguistics (pp. 136-144). Stroudsburg, PA: Association for Computational Linguistics.

Hunston, S., & Thompson, G. (2000). Evaluation: An introduction. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 1-27). New York: Oxford University Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 229–254). Amsterdam: Benjamins.

Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language, 87*(4), 771–816.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In P. Ladefoged, & V. Fromkin (Eds.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 115-132). Academic Press.

Labov, W. (1972). *Language in the inner city*. Philadelphia: University of Pennsylvania.

Linguistic Data Consortium. (2009). Broadcast conversation corpus produced for the DARPA Global Autonomous Language Exploitation (GALE) project. Audio used with permission.

Local, J., & Walker, G. (2008). Stance and affect in conversation: On the interplay of sequential and phonetic resources. *Text & Talk, 28*(6), 723–747.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge, MA: MIT Press.

Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-255). New York: Academic Press.

R Development Core Team. (2008). *R: A language and environment for statistical computing* (version 2.9.2) [computer software]. Retrieved from http://www.R-project.org.

Soltau, H., & Waibel, A. (2000). Specialized acoustic models for hyperarticulated speech. *Proceedings of the IEEE international conference on acoustics, speech, and signal processing (ICASSP '00)* (pp. 1779-1782).

Tomita, K. (2008). Effects of word familiarity in contexts on speaker's vowel articulation. *Bulletin of Yamagata University: Humanities, 16*(3), 55-67. Retrieved from http://www.lib.yamagata-u.ac.jp/kiyou/kiyouh/kiyouh-16-3/image/kiyouh-16-3-055to067.pdf.

Tomita, K. (2007). Effects of word frequency values on speakers' vowel articulation. *Bulletin of Yamagata University: Humanities, 16*(2), 65-75. Retrieved from http://repo.lib.yamagata-u.ac.jp/archive/kiyou/kiyouh/kiyouh-16-2-w065to075.pdf.

Uldall, E. (1960). Attitudinal meanings conveyed by intonation contours. *Language and Speech, 3*, 223-234.

Wennerstrom, A. K. (2001). *The music of everyday speech: Prosody and discourse analysis*. New York: Oxford University Press.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics, 23*(3), 349-366.

Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., & Iskarous, K. (2004). Vowel production and perception: Hyperarticulation without a hyperspace effect. *Language and Speech, 47*(2), 155-174.

Wichmann, A. (2002a). Attitudinal intonation and the inferential process. In B. Bel, & I. Marlien (Eds.), *Proceedings of the speech prosody 2002 conference* (pp. 11-16). Aix-en-Provence, France: Universite de Provence.

Wichmann, A. (2002b). Looking for attitudes in corpora. *Language and Computers, 40*, 247-260.